

Open Research Online

The Open University's repository of research publications and other research outputs

Creating a corpus of social workers' writing: Methodological challenges, representational issues and analytical concerns

Conference or Workshop Item

How to cite:

Leedham, Maria; Lillis, Theresa and Twiner, Alison (2018). Creating a corpus of social workers' writing: Methodological challenges, representational issues and analytical concerns. In: 4th Corpora and Discourse International Conference, 22-24 Jun 2018, Lancaster University.

For guidance on citations see [FAQs](#).

© [not recorded]



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Version of Record

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

Creating a corpus of social workers' writing: Methodological challenges, representational issues and analytical concerns

Maria Leedham, Theresa Lillis and Alison Twiner
The Open University, UK



Outline

1) Overview of WiSP research project

2) Building the WiSP corpus:

- methodological challenges
- representational issues
- analytical concerns

3) Some initial findings

The WiSP research project

- 3 year, ESRC-funded research project (Oct 2015- Oct 2018)
- The first national research project on writing in professional social work practice

Research Questions (for the whole project)

1. What are the institutional writing demands in contemporary social work?
2. What are the writing practices and perspectives of professional social workers?
3. What are the challenges faced and solutions found?
4. How are writing demands and practices shaping the nature of professional social work?

Builds on existing work

- Writing/literacies research (e.g. Barton and Hamilton; Barton and Papen; Lillis; Street),
- Everyday workplace literacies research (e.g. Brandt; Tusting; Smith)
- Management/recording systems and practices in social work (e.g. White et al; Huuskonen and Vakkari; Taylor)
- Professional discourse studies - oral (e.g. Roberts & Sarangi; Sarangi)
- Research on writing in social work (e.g. Pare; Rai and Lillis)

Why is this project important?

The production and use of written texts (often referred to as *paperwork*, *recording*, *inputting* or *documenting*) is a **high stakes activity** in professional social work.

Writing...

- plays a central role in all decisions about services for people
- is used to evaluate social workers' professional competence
- is often criticised – in public reviews and media reporting of high profile cases

Yet there is little systematic research on contemporary writing/recording demands, genres and practices....

Ethnography as our overarching epistemological orientation

There is no way in which language can be 'context-less' in this anthropological tradition in ethnography. To language, there is always a particular function, a concrete shape, a specific mode of operation, and an identifiable set of relations between singular acts of language and wider patterns of resources and their functions. (Blommaert, 2006, p. 4)

- Context-sensitive data collection methods and analysis in order to respond to research questions...
- A range of data...including interviews, observations, texts (case notes, emails, notes..), documentary data.
- **Corpus data is one of multiple data sets** within this ethnographic epistemological orientation.
- Aim of building the corpus is to assist in investigating SWs' written texts
- **Advisory panel of stakeholders:** social workers, SW managers, HEA, professional bodies e.g. HCPC

Data collection almost complete

- 5 agencies participating
- 70 social worker interviews
- 10 weeks of researcher observations
- 481 days of social worker writing activity logs
- 4,600 texts collected and anonymised
- 1 mw corpus
- *Still to do... screen capture of social worker at-desk writing*

Example texts

Single Assessment Part A - Your Family Household

Details of all children or young people living in household

Name	Date of Birth	Gender	Ethnicity	Disability	Religion	Who has PR	Party Record ID
[SU]	[DOB]	Unknown					[PIN]

Family or other household members

Name	Date of Birth	Gender	Ethnicity	Disability	Religion	Relationship to child or children
[PERSON1]	[DOB]	Female	British	None	Unknown	Mother

Other important people not living in household:

Name	Date of Birth	Gender	Ethnicity	Disability	Religion	Relationship to child or children	Address
[PERSON20]	[DOB]	Male	British	None	Unknown	Putative father	Unknown
[PERSON9]	[DOB]	Male	British	None	Unknown	Half-sibling	[LOCATION1]
[PERSON2]	[DOB]	Female	British	None	Unknown	Half-sibling	[LOCATION2]
[PERSON3]	[DOB]	Female	British	None	Unknown	Half-sibling	[LOCATION2]

Add Case Note

Title (max. 100 letters) *

Type of Note *

Date *

Time *

Note *

Add Alert Recipients for this case

Hello [PERSON1]

Please

Please

support

arrange

Is this

be further

commis

s.117 r

'open' pending this?

Hope this makes sense.

Thanks again,

[SW005] (Social Worker, [LOCATION1])

'[OA1]

twice th

[OA1]

this mo

they se

manag

stated

a weak

Contra

will hav

T/c message to [DSO] for advice on their capacity, t/c to [DAUGHTER] messages left X 2.

Spoke to [DSO] who rang to say she can provide the support hours and will get back to me with a start date.

I spoke to care-coordinator [PERSON3] – [OA1] who advised that the managers [OA1-MANAGER2] and [OA1-MANAGER1] where [sic] not in the office.

I advised that [SU008a] has 4 calls today with the names of 4 different carers, he usually has [CARER1] and [CARER2] which is the plan for the rest of the week

Methodological challenge 1: Gaining access to sensitive texts...



Methodological challenge 1: Gaining access to sensitive texts...

1. **Institutional** - formal research and governance procedures to seek permission from social services to carry out research
2. **Operational** - permission from service managers to approach social workers, depends on them teams having 'capacity' and assurance that research will not disrupt practice
3. **Individual** - agreement from individual social workers to take part (in different ways and levels) – some with payment

LA 1- (previous formal contact) level 1 secured at 6 and 12 months, insider contact ensured 2 and 3, **approved, strong**.

LA2- (previous formal contact) level 1 secured at 8 months, level 3 secured but failed because rejected at level 2. **rejected**.

LA3- (previous formal contact) additional legal agreement required around data protection, level 1 secured at 9 months, insider contact ensured 2 and 3, **approved, strong**.

LA4- (no previous formal contact) level 1 secured at 3 months, no support at level 2, slow development, **approved, weak**.

LA5- (no previous formal contact) and eventually rejected, renegotiated lower level of involvement secured at 12 months, level 3 some support, **approved, weak**.

LA6- (no previous formal contact) after several months we negotiated access to low level involvement, **approved, weak**.

Ethical issues....

Initial ethical approval from ESRC, OU, Local Authorities.... Ethics as an ongoing consideration for the 'virtuous researcher' (Oates)

But - Who should we ask for consent in using texts?

- Standard ethical procedure is to ask for consent of research participants. But who should we ask? - The Local Authority responsible for assigning the case? The social worker and writer of the text? Or the service user/client and focus of the text?

What we did

- gained consent from each participating Local Authority to access texts written by social workers – with proviso that texts were anonymised in-house and before we had access
- gained permission from the social worker-writers
- did *not* ask service users for their consent to use texts
- CF medical research (Mann et al, 2016)

Methodological challenge 2:

Anonymising the data

- Social workers write sensitive reports, often about vulnerable people, and often including personal or sensitive information

What is personal or sensitive data?

- *Personal data* relates to data that can identify a service user or other individual such as their home address or their date of birth.
- *Sensitive* personal data is information that has a greater risk of damage to the individual if it is misused or mishandled, such as criminal convictions, physical or mental conditions (Elliot et al, 2016).

How was anonymization carried out?

- Local Authority carried out the initial *redaction* of personal identifying features (name, GP name, etc) on site
- Asked redactors to replace personal details with *codes* e.g. [TEL], [CITY], [SU], [HEALTH], [SCHOOL], [CITY]
- Also... [SUD], [MUM EX-PARTNER], [BIRTH_MUM], [CHILD3]

Case note 1

Action - Ongoing Record

Phone call to [SCHOOL] school([TEL]) to arrange a direct work session to see [SU]. Visit arranged for the 2nd January at 11:30am.

Case note 2

Action - Ongoing Record

[SW201] Phone call to [MOTHERS-EX-PARTNER] (father to [HALF-SIBLING_2]). I spoke to [MOTHERS-EX-PARTNER] about the issues [SU-MOTHER] had raised regarding her contact with [HALF-SIBLING_2]. [MOTHERS-EX-PARTNER] said that [SU-MOTHER] can ring and speak to [HALF-SIBLING_2] anytime she wants. [MOTHERS-EX-PARTNER] said that issues was more that [HALF-SIBLING_2] has little interest in speaking to anyone on the phone as he would rather play and his interest has to be sparked on the phone. [MOTHERS-EX-PARTNER] said that he has not spoken to [SU-MOTHER] about contact so has not said that [HALF-SIBLING_2] can't go and stay and his view was that [SU MOTHER] is playing the martyr. [MOTHERS-EX-PARTNER] was clear that he wants [HALF-SIBLING_2] to have a relationship with [SU-MOTHER], and that she is more than welcome to pick him up and take him out for tea, but he suspects she will not do this as it requires her to make an effort.

Case note 3

Chronology

[DATE] - Incident with knife - No information relating to this found on [IT_SYSTEM].

[DATE] - Telephone call to CRU re. Concerns raised. Contact made with dad. [PERSON] had stopped taking meds. Dad states he does not require support at this time. NFA.

[DATE] - Referral to MAT. Mum taken significant overdose 4 x 28 [DRUG] Admitted to ITU.

[DATE] - Referral to MAT. Mum taken overdose 14 - 16 [DRUG] Admitted to hospital

[DATE] - Call from [HOSPITAL]. Mum transferred to CDU after taking an OD of anti-depressants and drinking a bottle of vodka.- Previous attendance at [LOCATION] Hospital.

[DATE] - Child referral.

[DATE] - No concerns for children in school. Being monitored.

[DATE] - [SU] exhibited sexualised behaviour by pulling a girls trousers down and asking her to do the same. Witness by Midday Supervisor in school.

[DATE] - Referral from school. Information that mum is having sexual relations with her brother. He has moved into the house.

Methodological challenge 3:

Archiving the corpus

Initial anonymization

Filename:

SW207_casenote_120315_3

I advised that [SU207a] has 4 calls today with the names of 4 different carers, he usually has [CARER1] and [CARER2] which is the plan for the rest of the week, [PERSON3] apologised and stated...

/matrix semtagging and

e medical conditions).

DATE]

ciated with texts

SW1 not [SW207]

Final anonymization

Filename: WISP1679

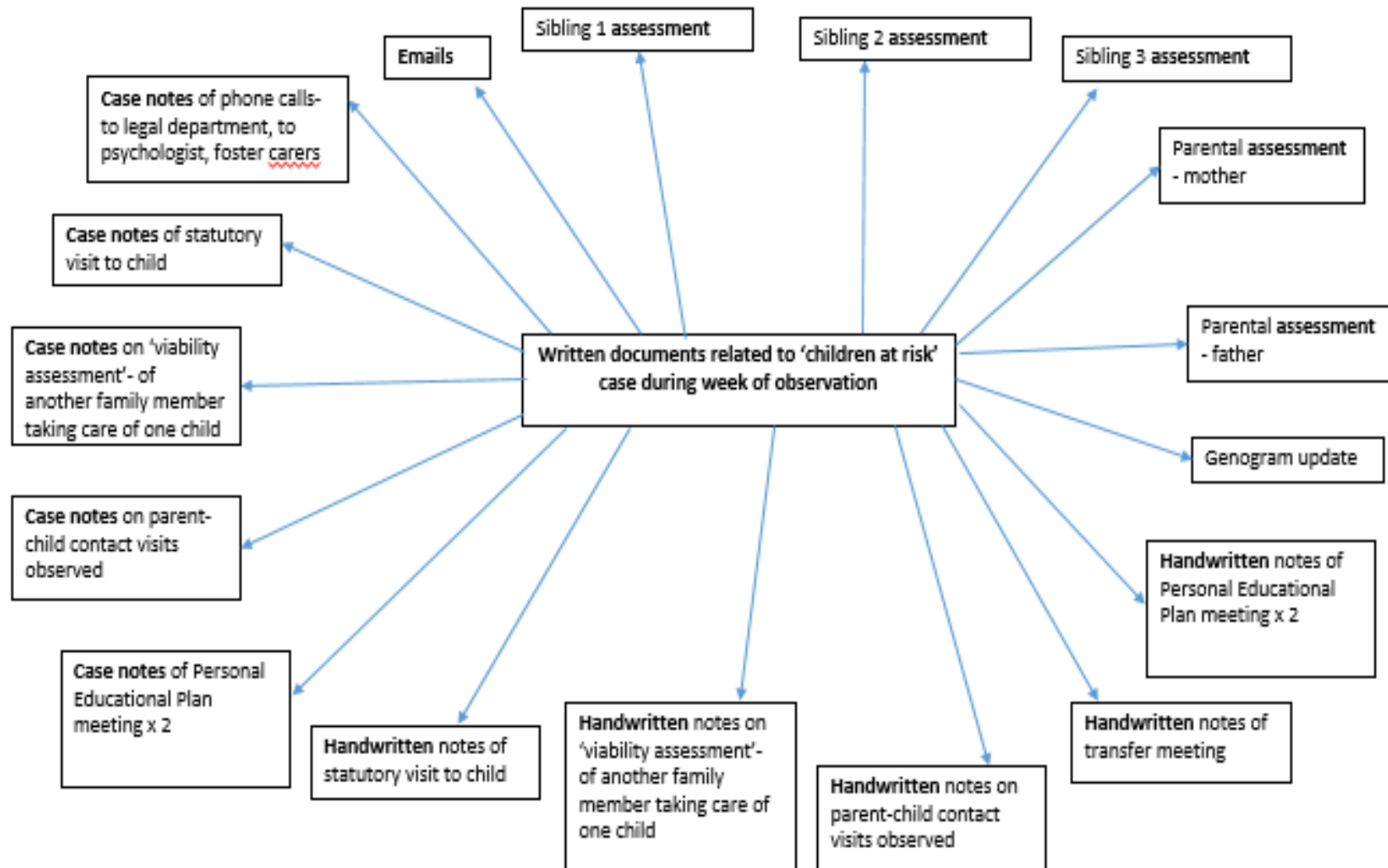
I advised that [SU] has 4 calls today with the names of 4 different carers, he usually has [CARER] and [CARER] which is the plan for the rest of the week, [PERSON] apologised and stated ...

- => Standalone texts rather
- Will restrict user access to
- Two corpora: team and arc

Representational issues

- How do we attempt to represent the writing of social workers?
- Sampling model? – across LAs, domains, text types, texts,
- Have a variety of social worker writers by age/ experience etc?
- All texts are not equal (update casenote vs critical incident casenote vs internal email vs court report)

Texts for a single case



Representational issues

- How do we attempt to represent the writing of social workers?
- Sampling model? – across LAs, domains, text types, texts,
- Have a variety of social worker writers by age/ experience etc?
- All texts are not equal (update casenote vs critical incident casenote vs internal email vs court report)
- Tag duplicates as don't want to over-represent any writer?.... But *what is a duplicate?* (non-SW writing? e.g. boilerplate disclaimers, form questions, VS. repeated SW language e.g. formulaic language, emails within casenotes, multiple assessments in one family, repeating action points....)

But...

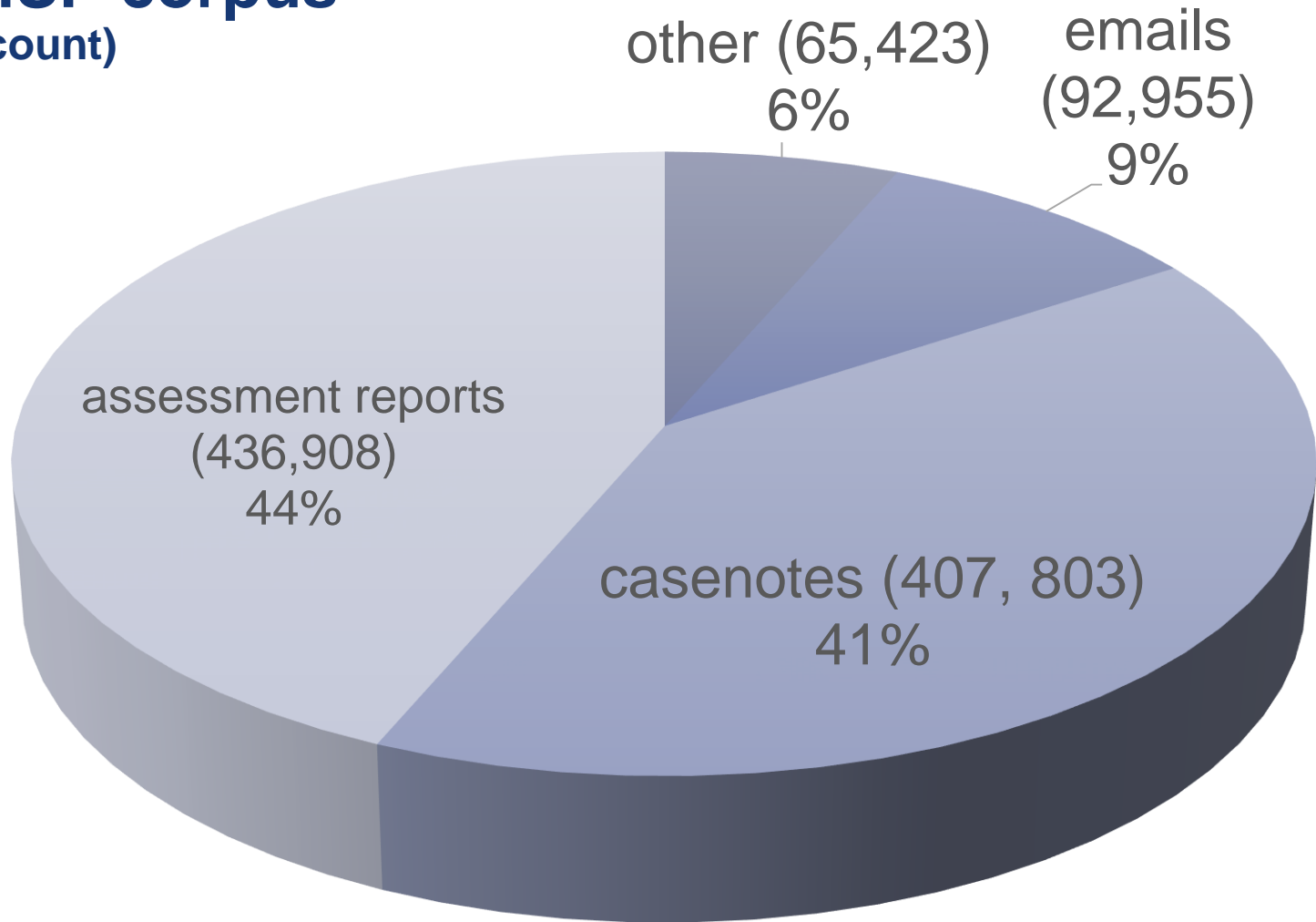
- Practical issues of access, time, money & goodwill

So... *take what you can get*

- And describe the corpus composition for future users
- Small corpus – 1 mw after deduping (from 1.2mw)
- NB Delinking of datasets and of writer from text makes the corpus less useful to other researchers

The WiSP corpus

(by wordcount)



Analytical issues... and some findings

- Wmatrix used to extract KW and MVA/Ls (Davies, 2000)

- BE06 as ref corp.

- Filtered on frequency

- Filtered on frequency

- Filtered on frequency

- Candidate

- CKIs checked to

across all 3 SW

No	Category	Example key items
1	SW roles and activities	care*, LAC, proceedings, safeguarding,
3	family & institutionalised care	adopters, birth_mother, care*, contact*, foster, unborn,
5	Primarily evaluating	abusive, appropriately, behaviours, can't, concerns, due_to, enjoys, exploitation, misuse routines, struggles,

Our example: *They shared all evening meals together.*

= evaluative language?

She needs to have better routines and ensure...

= evaluative within SW domain of children

- Asked advisory

[PERSON] continues to misuse heroin.

- Emphasises illegal drug use/excessive alcohol use

processing, assault, attends, bail, belongings, quising, home_environment, meals, siding, toilet, unwell, utilise,

abusive, appropriately, behaviours, can't, concerns, due_to, enjoys, exploitation, misuse routines, struggles,

d, contacting, ed, hi, info,

Comparing text types

No.	Category	Key items		
		Case notes	Emails	Assessment reports
4	<i>Primarily describing a situation</i>	<i>arrived asleep bail</i>	<i>finance allowance funded</i>	<i>achieve appearance at_risk</i>
5	<i>Primarily evaluating</i>	<i>attempted shouted worried</i>		<i>alleged befriends illicit</i>
6	<i>Communication</i>	<i>advised discussed suggested</i>	<i>attached catch_up copies</i>	<i>caller</i>

- Is it possible to separate ‘describing’ and ‘evaluating’?
- Is all SW writing evaluative?

The example of *asleep* – evaluative?

- [CHILD] looked comfortable with [MUM] and he fell **asleep** on her. He appeared content in the home and [MUM] was attentive to his needs. She fed him during the visit. (visit to toddler and parent)
- She is not keeping him clean, there was no fresh food at home. His bed was not made and all rooms were very dirty, untidy and smelly. [CHILD] said he sometimes slept on the sofa with Mum and fell **asleep** watching TV with her. The home conditions and parenting are not good enough at present. (statutory visit to a 10 year old child)
- [CHILD] stated he started a paper round to get extra money for himself however [CHILD] would fall **asleep** at school.
- We knocked at the door several times before we got an answer. [DAD] then came to the door and appeared very sleepy as though he had just been woken up. [CHILD] was wondering about and appeared happy. [DAD] invited us in. I asked [DAD] if he had been **asleep** and he stated he was very tired from doing his night shift, he stated he hadn't been **asleep** but his eyes had been shutting but he could still hear everything.
- [SU] still unsteady mobilising and spending a lot of time **asleep**. (casenote, adults)

Comparing text types

1. Told [CHILD] to apologise, *shouted* "for fucks sake...."
2. her with not going on holiday again, *shouted* 'no!' took her by the hand
3. Do you want to apologise for your behaviour?' *shouted* 'No!' then said
4. shouted from the lounge 'mum' *shouted* back to him 'if you want me come
5. go and see if [PERSON] wants to play?' *shouted* 'No' slammed the cupboard doors

[PERSON] *stated* that she is not sure how [PERSON] ([SU]'s son) will feel about this. She *asked* "do you think that they will think I am awful for not having him back". I *advised* that I have spoken with [PERSON] (Son) and he was open to the possibility of [SU] remaining in residential care. I *advised* [PERSON] that she has to be honest how she feels as [SU] could only return home if she is able and willing to care for him.

recent or forthcoming events (*11am, last_night*)

- Report in past tense on arrangements
- Different reporting of SW and SU talk (*advised, discussed, agreed, shouted*) – quotations to provide evidence

Conclusions aka *What I've learned...*

- Don't underestimate the issues involved in compiling a corpus of sensitive texts!
- Stakeholders and participants can offer valuable insights
- Important to try out ways of slicing the data to ensure findings are robust
- Need to consider variables within the data
- Don't over-promise what you can archive...

References

- Barton, D. & Papen, U. (2010). *The anthropology of writing. Understanding textually mediated worlds*. London.
- Brandt, D. (2009). When people write for pay, *Journal of Advanced Composition* 29:1, 2: 165-197.
- Gabrielatos, C., 2018. "Keyness Analysis: nature, metrics and techniques". In C. Taylor and A. Marchi (eds.) *Corpus Approaches to Discourse: A critical review*. Chapter 12. Oxford: Routledge.
- Lillis, T., Leedham, M. and Twiner, A. (under review) "'If it's not written down it didn't happen': contemporary writing and recording demands in social work". For *Journal of Applied Linguistics and Professional Practice*.
- Lillis, T. (2013). *The sociolinguistics of writing*. Edinburgh: EUP
- Paré, A. & Le Maistre, C. (2006). Active learning in the workplace: Transforming individuals and institutions. *Journal of Education and Work* 19:4: 363-381.
- Rai, L. & Lillis, T.. (2012). 'Getting it Write' in social work: exploring the value of writing in academia to writing for professional practice. *Teaching in Higher Education*, 153/4. 5-70.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins.
- Elliot, M., Mackey, E., O'Hara, K. & Tudor, C. (2016). [The Anonymisation Decision-Making Framework](http://ukanon.net/wp-content/uploads/2015/05/The-Anonymisation-Decision-making-Framework.pdf). Manchester: UKAN. <http://ukanon.net/wp-content/uploads/2015/05/The-Anonymisation-Decision-making-Framework.pdf>. Accessed on 190618.
- Kubanyiova, M. (2008). Rethinking research ethics in contemporary applied linguistics: the tension between macroethical and microethical perspectives in situated research. *Modern Language Journal*, 92(4): 503 - 18.
- Mann, S.P., Savulescu, J. & Sahakian, B. (2016). Facilitating the ethical use of health data for the benefit of society: Electronic health records, consent and the duty of easy rescue. *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, 374 (2083), 1-17.
- Rier, D.A. (2004). Publication visibility of sensitive public health data: When scientists bury their results. *Science & Engineering Ethics*, 10 (4), 597-613.
- Rayson, P. 2008. "From key words to key semantic domains". *International Journal of Corpus Linguistics*. 13:4 pp. 519-549. DOI: 10.1075/ijcl.13.4.06ray

Thank you

@OUWISP #OUWISP

www.writinginsocialwork.com